

CSC 447 - Artificial Neural Networks

Stephanie Athow, Ben Kaiser, Marcus Haberling

February 24, 2016

1 Training results

How well does your network train?

The following results for this section were obtained using a neural network topology of 25 input nodes, 3 or 4 hidden nodes (dependent on which parameter file was used) and 3 output nodes. The momentum term was set to 0.8 and the learning rate was set to 0.4.

This network topology appears to train really well. With 1-2 training sessions, the RMS values are often around 0.13 for both the Black Hills data and the Northwest South Dakota data.

When training the network with the Black Hills data, if the same Black Hills data is run as a test, the accuracy we got was between 94% and 100%. However when the Northwest data was run on the same net, the accuracy was lower, usually between 75% and 81.25%.

When the network was trained on the Northwest data, if the same Northwest data was run as a test, the accuracy was high, several times reaching 100%. However, the Black Hills data does not perform very well with accuracy down near 50%.

If the network was trained on both data sets, we noticed that order did seem to matter. If Black Hills was trained first and Northwest trained second, testing the Black Hills data we got results between 69-75%. The Northwest data set would give results between 80-100%.

If the Northwest data was trained first and the Black Hills data was trained second, the results were not as good. Black Hills results were about the same but the Northwest data was down around 60-80%.

During cross-validation, we generally got 100% which is a little surprising. When the network was trained on the Black Hills data but the weights file was not deleted and the network was trained again on the Northwest data, we got poor results (35-80%) which is not unexpected.

2 Network Topology

What is the impact of network topology (i.e., changing the number of hidden layer nodes) on training?

One noticeable trend was when the number of hidden layers increased, the training rate decreased but would usually converge. Generally, more than 3 layers of hidden layers would take many more training epochs for the RMS values to decrease. An example topology of this would be (25 4 4 4 3). However, topologies with a large number of middle layers had a very hard time converging as the rate nearly stopped.

In addition to many hidden layers slowing RMS convergence, more nodes in the hidden layers mean the network took longer to train. An example topology of this is (25 200 3). Interestingly enough, a topology of (25 25 3) performed comparably to the (25 4 3) topology.

Additionally, not all network topologies trained well, a topology of (25 7 7 7 7 7 7 3) would not train to an RMS below 0.44.

3 Generalization of Data

How well does the network generalize from training data to testing data?

The network does not generalize well between training data and testing data. If the network was trained on the Black Hills data, testing the Black Hills data on it could be 80-100%. However, if the testing data was switched to the Northwest set, accuracy could drop to 50% and the same for the reverse situation. When the training involved both sets of data, the Black Hills data dropped in accuracy (60-80%) but the Northwest data increased in accuracy (60-80%).